Int J Appl Earth Obs Geoinformation 78 (2019) 261–267

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Hourly gridded air temperatures of South Africa derived from MSG SEVIRI

Hanna Meyer[a,*], Johannes Schmidt[b], Florian Detsch[b], Thomas Nauss[b]

[a] Institute of Geoinformatics, Westfälische Wilhelms-Universität Münster, Heisenbergstr. 2, 48149 Münster, Germany
[b] Faculty of Geography, Philipps-Universität Marburg, Deutschhausstr. 10, 35037 Marburg, Germany

ARTICLE INFO

Keywords:
Air temperature
Climate
Machine learning
Meteosat
Random Forest
South Africa

ABSTRACT

Monitoring of climate variables such as air temperature is gaining increasing importance under climate change. This study aimed at developing an hourly gridded 3.5 × 3.5 km air temperature (T_{air}) data set for entire South Africa. In a Random Forest approach, MSG SEVIRI data from 2010 to 2014 were used and related to T_{air} measured by 78 weather stations. An external validation on new climate stations and years that were not used for model training indicated the ability of the model to predict T_{air} with a RMSE of 2.61 °C and a R^2 of 0.89. The resulting model can be applied to the entire MSG SEVIRI time series since 2004. It hence allows for spatio-temporal pattern analysis as well as for the detection of trends which is relevant in the context of climate change.

1. Introduction

South Africa's climate is predicted to undergo considerable changes (Niang et al., 2014; Ziervogel et al., 2014; Jury, 2013) but at the same time, the area is considered to be highly vulnerable to climate change (Niang et al., 2014; Ziervogel et al., 2014). This raises the need for a spatially explicit monitoring of climatic variables to assess current patterns and trends. Amongst climatic variables, air temperature (T_{air}) is the most striking key-parameter of climate change and a crucial parameter for ecosystem processes. Its monitoring is hence an important task. T_{air} in South Africa (as in other parts of the world) is traditionally monitored on the basis of weather station networks (Kruger and Sekele, 2013). However, a sparse network of stations hardly allows studying spatio-temporal dynamics since temperature patterns vary considerably in space (Kruger and Sekele, 2013). Spatial interpolations of weather stations, such as performed by Eiselt et al. (2017) are frequently used to fill the gaps between stations, however, spatial interpolations as well require a sufficiently dense network of weather stations which is not given for entire South Africa. Remote sensing methods are therefore indispensable and have large potential especially in areas with a comparably low cloud frequency such as in South Africa.

To derive T_{air} from remotely sensed data predictive modelling is a frequently used method where statistical models are set up between satellite information and measured T_{air}. In this context, different remote sensing data have been used as proxy for air temperature. Prihodko and Goward (1997) showed that the satellite derived NDVI serves as a strong predictor for air temperature. Even more obvious predictors can be derived from the thermal bands of different satellite sensors. Especially data from the Moderate Resolution Imaging Spectroradiometer (MODIS) find frequent application to estimate T_{air} on a comparably high spatial resolution of 1 km. Many studies used the MODIS land surface temperature (LST) products as remotely sensed predictors in different regions of the world to estimate T_{air} in a spatio-temporal way (e.g. Janatian et al., 2017; Xu et al., 2018; Meyer et al., 2016; Vancutsem et al., 2010; Shi et al., 2016; Kilibarda et al., 2014; Neteler, 2010). The data have been used to estimate air temperature from very local scales such as for the analysis of urban heat islands (Huang et al., 2017b) to global scales (Hooker et al., 2018). Though LST is physically different from T_{air}, it is a strong proxy for the energy a surface receives and serves in relation with other variables such as information on the vegetation or land cover as a strong predictor for T_{air}. MODIS data offer four overflights per day (two of Terra and two of Aqua, more towards the poles) which causes the data to be limited for estimating air temperature on a sub-daily basis. Despite having a lower spatial resolution, sensors onboard geostationary satellites such as the Meteosat Second Generation Spinning Enhanced Visible Infrared Imager (MSG SEVIRI), in contrast, have the advantage of a higher temporal resolution (15 min) and provide temporally more consistent data in similar spectral ranges. MSG SEVIRI and comparable geostationary satellite data have found frequent application as predictors for other climatic parameters as e.g. rainfall (Kühnlein et al., 2014; Giannakos and Feidas, 2013; Meyer et al., 2017) or cloud properties (Benas et al., 2017). However, also the suitability for operational monitoring of T_{air} is indicated (Stisen et al., 2007; Nieto et al., 2011; Good, 2015). Recently Good (2015) used MSG SEVIRI derived LST data to estimate daily T_{air}

* Corresponding author.
E-mail address: hanna.meyer@uni-muenster.de (H. Meyer).

https://doi.org/10.1016/j.jag.2019.02.006
Received 8 January 2019; Received in revised form 14 February 2019; Accepted 14 February 2019
Available online 26 February 2019

over Europe with very promising results. But recent approaches only focused on the thermal channels of MSG SEVIRI and related them to measured T_{air} via regression analysis. Using only infrared channels, however, ignores dynamic land surface characteristics influencing T_{air}. The spectral resolution of MSG SEVIRI ranges from visible and near infrared to water vapour and thermal infrared channels, serving as proxies for not only surface temperature, but also vegetation characteristics which can be regarded as strong predictors for T_{air}. While a majority of the studies that use remotely sensed data to estimate T_{air} rely on parametric modelling approaches such as linear models (Huang et al., 2017a; Benali et al., 2012), machine learning algorithms, such as the well known Random Forest algorithm, have been shown to be superior in making accurate predictions (Meyer et al., 2016; Noi et al., 2017) which is important when diverse predictor variables such as the different information from MSG SEVIRI are being used.

In this study, the comprehensive information of the MSG SEVIRI data is used in a machine learning-based approach to estimate spatially continuous hourly T_{air} of South Africa. The aim is to train a model that can be applied to the entire time series of MSG SEVIRI, allowing the spatio-temporal monitoring of hourly T_{air} of South Africa since 2002 and in approx. 3.5 km spatial resolution.

2. Methods

2.1. Study area

The area of investigation comprises entire South Africa, including Lesotho and Swaziland (Fig. 1). The terrain raises from the coastal lowlands to approx. 1500 m on the Central Plateau. Average annual rainfall in South Africa follows an aridity gradient from the dry west to the more humid east (Kruger, 2007). In the summer months (Dec-Feb) warmest temperatures are in the arid north west which features a hot desert climate (Kruger, 2004). Here, average summer temperatures can be above 35 °C. On the Central Plateau the climate is semi-arid with cool winter temperatures. The east features a subtropical climate. Savannas and grasslands are the dominant land cover types in the north and north-east while the central western part of South Africa can be assigned to the Nama-Karoo type. The cape region is covered by the Fynbos vegetation (Mucina and Rutherford, 2006).

2.2. Data and pre-processing

2.2.1. Ground truth data

Quality controlled reference data were obtained from the South African Weather Service (SAWS). In total, data from 78 stations for the years 2010 to 2014 were considered (Fig. 1). All stations provided T_{air} records in hourly or sub-hourly temporal resolution. For the purpose of a unique processing scheme, all data were averaged to hourly T_{air}. Overall measured hourly T_{air} was 17 °C in average with a standard deviation of 7.5 °C and a minimum of −11 °C and maximum of 47 °C. Though the station data are not randomly distributed in the model domain, they cover a large temperature gradient, from sites with comparably cold temperatures (~10 °C), to sites in areas with highest (> 20 °C) yearly means of T_{air} (Fig. 1).

2.2.2. MSG SEVIRI channels as spectral predictors

MSG SEVIRI channels (Aminou et al., 1997) were used as satellite-based predictors for hourly T_{air}. MSG SEVIRI scans the full disk every 15 minutes with a spatial resolution of 3 × 3 km at sub-satellite point (approx. 3.5 × 3.5 km in South Africa). Reflected and emitted radiances are measured by three spectral channels at visible (VIS) and very near infrared (NIR, between 0.6 and 1.6 µm), eight channels ranging from near-infrared to thermal infrared (IR, between 3.9 and 14 µm) and one high-resolution VIS channel.

MSG SEVIRI was obtained as level 1.5 data (EUMETSAT, 2010) and preprocessed to radiances according to EUMETSAT (2012a) and brightness temperatures according to EUMETSAT (2012b) using a processing scheme based on a custom raster processing extension of the eXtensible and fleXible Java library (see https://github.com/umr-dbs/xxl) which enables parallel raster processing on CPUs and GPUs using OpenCL. The processed MSG SEVIRI scenes, as well as the corresponding solar zenith angle information, were aggregated from 15 minutes to one hour (median aggregation) to match the temporal resolution of all available weather station data. To ensure a high data quality, only pixels where all four scenes of an hour were cloud-free were taken into account for further analysis.

All MSG SEVIRI channels except for the high-resolution VIS were used in this study. The data provide very intuitive proxies for T_{air} via its thermal infrared channels, but also allow to include vegetation characteristics via its VIS and NIR channels. The corresponding solar zenith angle of the MSG SEVIRI data further includes information about seasonality and time of the day into the model. Note that the VIS channels have near-zero reflectance values during nighttime conditions, hence during model training the algorithm will have to rely on the other channels during nighttime.

2.2.3. MODIS NDVI 8 day composites

As a proxy for vegetation characteristics, the 16-day MODIS NDVI products MOD13Q1 (Didan, 2015a) and MYD13Q1 (Didan, 2015b) collection 6 were used and merged to one product that reflects the NDVI

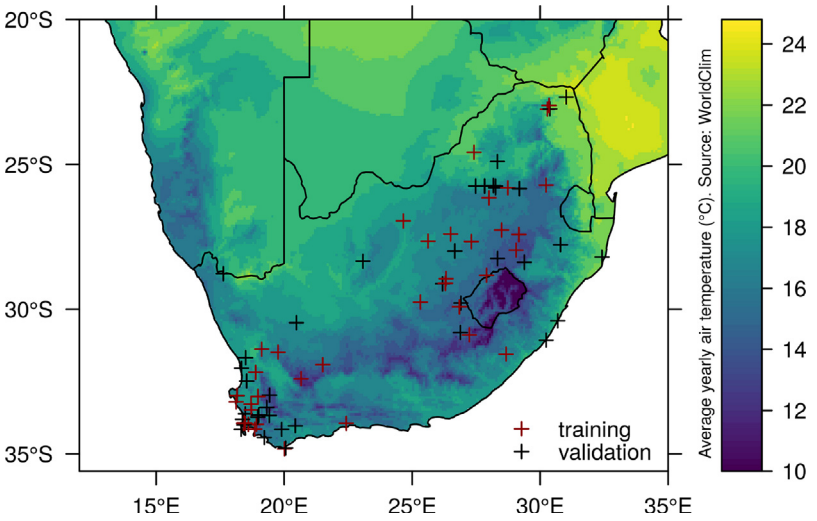

Fig. 1. Map of the average T_{air}s in the study area as estimated by WordClim (Hijmans et al., 2005). Symbols show the locations of the weather stations that were used as ground truth data in this study. Red symbols indicate stations used for model training, black symbols indicate stations used for model validation. Source of the weather stations is the South African Weather Service (SAWS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as an 8-day product. Though NDVI can be calculated from MSG SEVIRI as well, the MODIS 8-day composite was included as further potential predictor as it provides stable information over 8 days that were applied for daytime as well as nighttime conditions.

2.2.4. Cloud mask

As surface temperature for MODIS LST cannot be given for cloudy conditions, a cloud mask was used to restrict the analysis to cloud-free MSG SEVIRI pixels. For 2010 to 2012, the CM SAF CMa Cloudmask product (Kniffka et al., 2014) was applied. Due to the availability at the time of data acquisition two different cloud mask products were used: the CM SAF CMa cloudmask was used for the years 2010–2012, and the comparable CLAAS-2 mask (Finkensieper et al., 2016) was applied for the years 2013 and 2014. CLAAS-2 is the 2nd edition of the SEVIRI-based cloud property data record provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF; see also Stengel et al., 2014 for further information). It is now available for the entire MSG SEVIRI period since 2004 (Benas et al., 2017).

2.2.5. Auxiliary static variables

Beside of MSG SEVIRI data and the corresponding solar zenith information, four temporally static potential predictors were included: elevation, continentality, the MSG SEVIRI satellite viewing angle and the month of the major rain season for each pixel in the model domain. Elevation data were retrieved from the SRTM mission (Jarvis et al., 2008). Continentality was derived as distance from the sea and the timing of the rainy season was derived from WorldClim (Hijmans et al., 2005) monthly long-term averages of precipitation.

2.2.6. Overview on the predictor variables

As potential predictors, all MSG SEVIRI information including the solar zenith angle, MODIS NDVI, as well as the auxiliary static variables were included. Fig. 2 shows the linear correlations between the predictor variables in the dataset. It also indicates that the maximum individual linear correlation to measured T_{air} was reached for the IR channel centered at 10.8 µm with a correlation coefficient of 0.79 (R^2 = 0.62).

2.3. Random Forest modelling

Random Forest (Breiman, 2001) was used to learn the relationships between the predictors and measured T_{air}. Random Forest was chosen due to its ability to deal with numeric (MSG SEVIRI) as well as categorical (rainy season) information. It has also shown to have great potential in comparable studies (Appelhans et al., 2015).

The Random Forest modelling procedure consisted of data splitting in order to create independent training and test data, variable selection, model tuning (mtry) and the final spatial prediction. For all models, the number of trees (ntree) was kept constant at 500, after no increase in performance could be observed using a higher number of trees. The model was validated for its ability to make predictions on new locations as well as new points in time by use of independent data. These steps will be described in detail in the following.

2.3.1. Compilation of training and test data

The predictor variables were extracted for the locations of the climate stations and merged with the measured air temperature at the respective time of observation. For the NDVI which represents a 8 day composite, the NDVI values which were temporally closest to the air temperature measurement were considered. The extracted and merged dataset then consisted of 2,000,000 matching and cloud-free data points.

The weather stations were randomly split into stations used for model training and stations used for model testing. The time series of 40 stations were used for model training, the remaining 38 stations for external model validation (Fig. 1). From these 40 stations, only the years 2010 and 2011 were used for model training, and the years 2012–2014 were used for the external model validation. Using these splitting criteria, the final training data set contained 430,000 data points and the external validation set contained 630,000 data points.

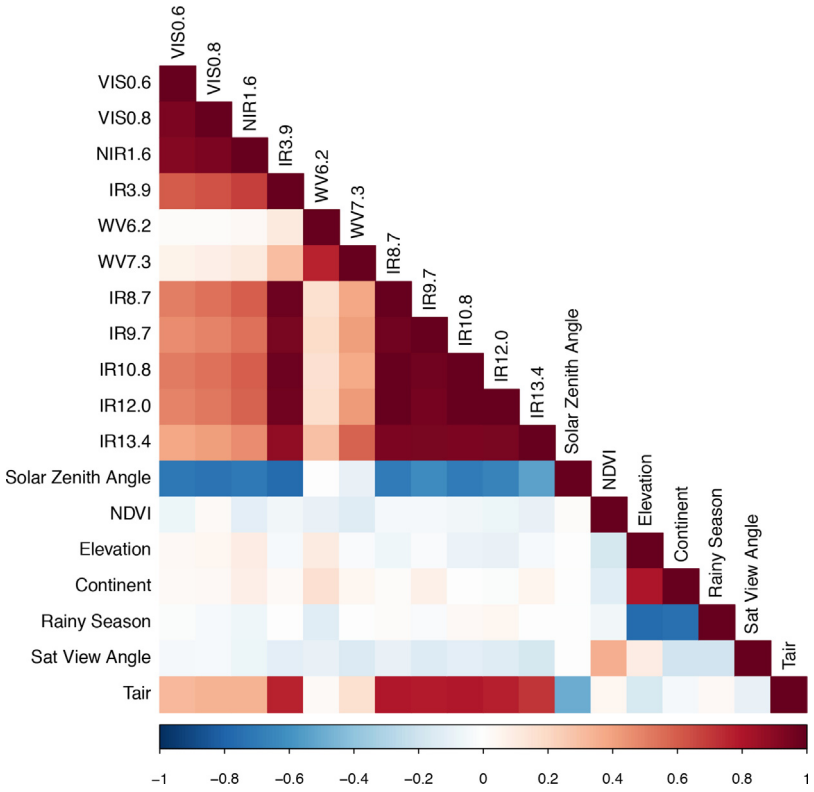

Fig. 2. Pearson correlation between potential predictor variables and measured air temperature.

2.3.2. Model tuning, training and variable selection

The requirement on the model was its ability to make predictions for new time steps and beyond the spatial locations of the training stations. Therefore, modelling strategies were chosen in view to this target. The optimal tuning parameter mtry (for further explanations see Kuhn and Johnson, 2013) was assessed using 10-fold "Leave-Location-and-Time-Out" (LLTO) cross-validation (CV) (Meyer et al., 2018). Hence the data were split into 10 folds, each fold consisting of the data of 1/10 of the days available in the training data set and 1/10 of the weather stations. Ten models were then trained each leaving the data of one fold out and the performance was estimated using the respective held-back data. LLTO CV was not only used to determine the optimal tuning parameters but also to assess the importance of the potential predictor variables in view to their relevance to predict T_{air} beyond the location of the training stations and beyond the training years. Therefore, a forward feature selection (FFS) in conjunction with LLTO CV as implemented in the CAST package (Meyer, 2018) was applied, that searches for the best combination of variables with "best" refers to the lowest error in view to left out weather stations and left out days. To do this, the FFS first tests all possible combinations of two predictor variables in a Random Forest model and assesses the model performance using LLTO CV. Based on the best performing two variables, the number of variables is increased. The algorithm stops, when none of the remaining predictor variables can increase the LLTO CV performance. See Meyer et al. (2018) for further description of this variable selection method and its relevance for spatio-temporal machine learning applications. As the FFS is very computation time consuming, only 20,000 data points from the training data set were considered to select the optimal variables. They were determined by stratified random sampling which kept the distribution of T_{air} equal to the original full training dataset. Once the optimal variables were determined, the final model was trained using a subset of 150,000 data points which were again chosen by stratified random sampling. Mtry was extensively tuned for the final model between 2 and the number of predictor variables using LLTO CV.

2.3.3. External validation

Though the target-oriented CV gives the error in view to new locations and new points in time, it is advisable to use completely independent data for model validation if the amount of data allows for that. Therefore, the data of 38 stations from the years 2012, 2013, and 2014 were used for an external model validation. They can be seen as completely independent as neither any measuring values from the years 2012–2014 nor any measurement of the respective 40 stations have been used for model training. R^2 and RMSE were used as validation metrics to assess the model performance in view to these new data.

3. Results

3.1. Model performance and contribution of the predictor variables

The FFS indicated that the optimal two variables were the infrared channel centered at 3.9 µm and the sun zenith angle. Using these two variables, T_{air} could be predicted with a RMSE of 3.72 °C. The further inclusion of NIR channel centered at 1.6 µm could decrease the RMSE to 3.46 °C. The optimal model could be retrieved by further including the variables elevation, IR at 12.0 µm, IR at 10.8 µm, IR at 8.7 µm, IR at 13.4 µm, IR at 9.7 µm, the reflection in the visible spectrum at 0.8 µm, and the water vapour band centered at 7.3 µm (importance in decreasing order). With these 11 variables, T_{air} could be modelled with a RMSE of 2.75 °C and a R^2 of 0.86. Including any 12th variable increased the LLTO CV error (Fig. 3). These results of the optimal model are comparable to the outcome of the external validation where complete years and climate stations were left out (R^2 = 0.89, RMSE = 2.61 °C, Fig. 4). The model slightly underestimated warm temperatures and overestimated cold temperatures, thus very warm or cold extreme values were not predicted by the model. The performance did not show any obvious temporal patterns or trends neither over the years, nor in seasonality.

3.2. Spatio-temporal T_{air} predictions

The application of the trained model to the time series of the MSG SEVIRI data allowed for spatial predictions with the performance mentioned above. Fig. 5 shows as an example the spatial T_{air} patterns for the year 2013 aggregated to a monthly basis.

4. Discussions

The results have shown great potential in using MSG SEVIRI and machine learning algorithms for the monitoring of T_{air}. The choice of MSG SEVIRI as baseline data in contrast to frequently used MODIS LST products allows for the high temporal resolution of hourly data and a more continuous data set. Hence the high temporal resolution allows for temporally small-scale patterns in T_{air}. It further allows for the delineation of minimum and maximum or mean daily temperature which are frequently used T_{air} summaries which are problematic using MODIS LST due to only four overflights. However, the approach has the drawback of a lower spatial resolution so that in future approaches it might be an idea to combine MODIS LST and MSG SEVIRI data to obtain high resolution T_{air} data sets in both spatial and temporal dimension.

From a methodological modelling perspective, the need for FFS with target-oriented validation became obvious to assess the importance of variables to make predictions in space and time. If this is not done there is a risk that unsuitable variables act in a counterproductive way and decrease the performance of the models. In this context, the selection of the variable elevation by the FFS comes a bit as a surprise as Meyer et al. (2018) have shown that temporally static variables are prone to overfitting in spatio-temporal prediction tasks. However, this was attributed to the fact that by the choice of predictor variables the algorithm was given access to the individual time series by a combination of unique temporally static variables but also temporal variables like the day of the year. This was avoided in this study because temporal variables were substituted by the solar elevation that gave an indication about potential solar irradiation in a spatio-temporally dynamic way. Further surprising was that the NDVI was not selected by the FFS as an important variable. It has shown to be important for the difference between LST and T_{air} (Vadász, 1994). In this study NDVI had no individual linear relationship to T_{air} as indicated in Fig. 2. However, if the relationship is analyzed for each weather station individually, clear spatial patterns can be observed with positive correlations in the north east and negative correlations in the South West of South Africa (Fig. 6). This pattern can be explained by the time of the rainy seasons in South Africa where the South West receives most rainfall in winter and the North east receiving most rainfall in summer (Kruger, 2007). Though it was accounted for this in the model by including the season of the rainy season as auxiliary predictor variable, the NDVI could not improve the T_{air} predictions. Apparently the MODIS NDVI did not add additional value to the MSG SEVIRI bands.

A comparison between the results of the cross validation to the results of an external validation indicated that the model can be applied to new MSG SEVIRI scenes and beyond the locations used for model training as no overfitting could be observed. From a methodological perspective it could be shown that using large data sets as applied in this study, a target-oriented cross validation is sufficient and a retention of an external validation set is not crucial as no significant deviations between the two validation strategies could be observed.

The performance of the presented model is similar or better than other studies that use mostly MODIS LST to estimate air temperature Zhu et al. (2013), Huang et al. (2017a), Benali et al. (2012), Shi et al. (2016), Hengl et al. (2012), despite the strict validation strategy used in

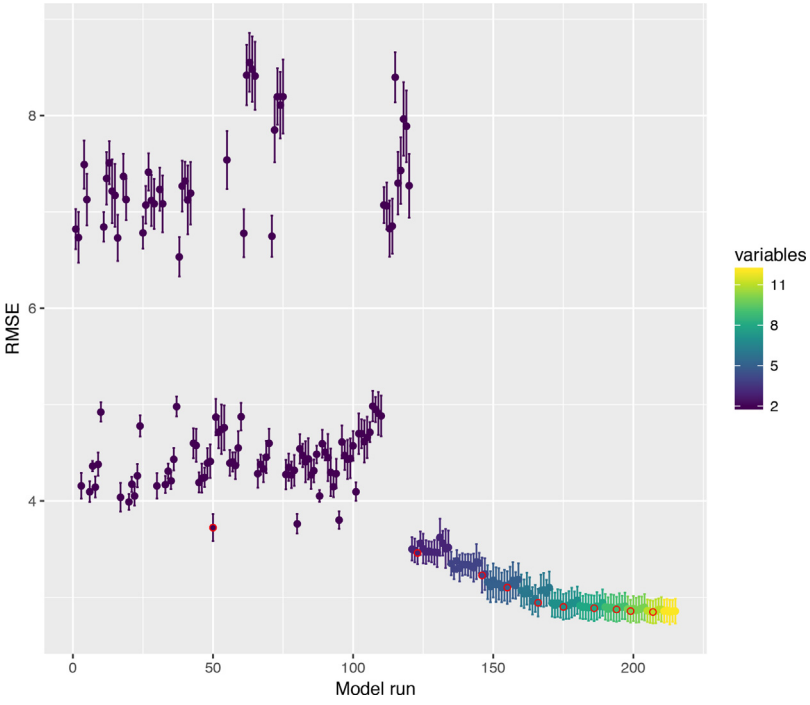

Fig. 3. Results of the forward feature selection. The mean RMSE and its standard deviation over a 10-fold "Leave-Location-and-Time-Out cross-validation are shown for each model run. Colors visualize the number of variables used in each run. The best models for each number of variables are marked with a red circle. The addition of any 12th variable (shown in yellow color) did not improve the model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

this study and the high temporal resolution of one hour. Differences in the validation approach, however, do not allow for a direct comparison. Most other validation studies do not rely on spatio-temporal validation. Meyer et al. (2016) have shown that the performance estimation can differ considerably depending on the validation strategy being used. However, since temporal and spatial components are held back for validation, the performance estimates in this study represent a more unbiased estimation compared to studies that base on random validation (Meyer et al., 2018).

The limitations of the developed T_{air} data set is the prediction of extreme values. This is a common problem for ensemble techniques such as the used Random Forest algorithm (e.g. see Elith and Graham, 2009), however it seamed reasonable due to the high overall model performance. In view to changing climate, recurrent re-training might be an option to adapt the model when temperatures generally raise.

The resulting T_{air} data set can serve as a baseline in trend and pattern analysis and ongoing monitoring of T_{air} with climate change. The temporal validation indicated that it can be used beyond the years used for model training and the spatial validation indicated that the model can be applied on the model domain of South Africa. An exception are the Drakensberges that feature distinct temperatures but no weather stations could be included in the model training in this study so that no reliable estimates are possible and the results for this area should hence be regarded with care.

The results have shown that the satellite-based monitoring of air temperature allows for accurate predictions and a comparable high

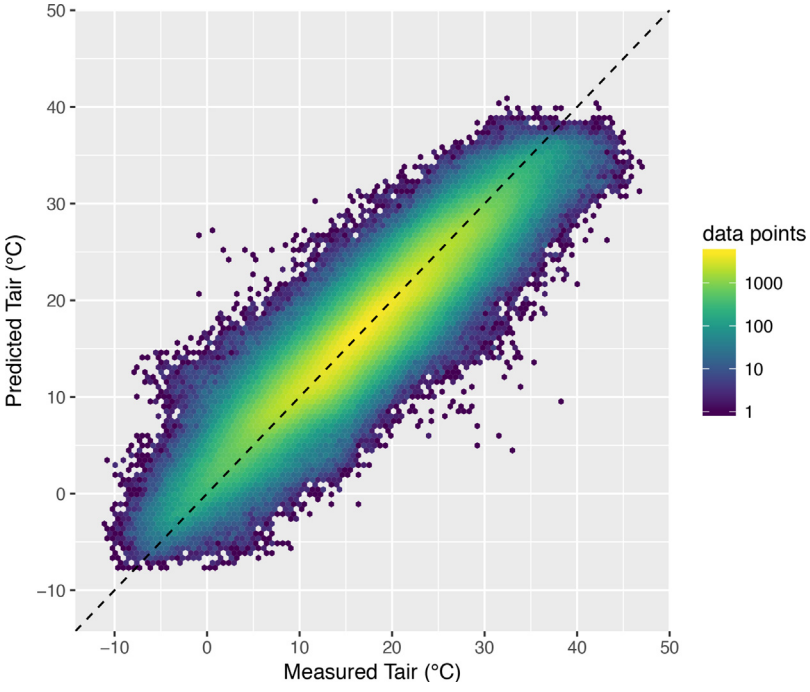

Fig. 4. Comparison of estimated hourly air temperatures quantities for the testing years 2012–2014 to measured air temperature by 38 weather stations that have not been used for model training. For an easy visual interpretation, the data are presented via hexagon binning in which the number of data points falling in each hexagon is depicted by color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

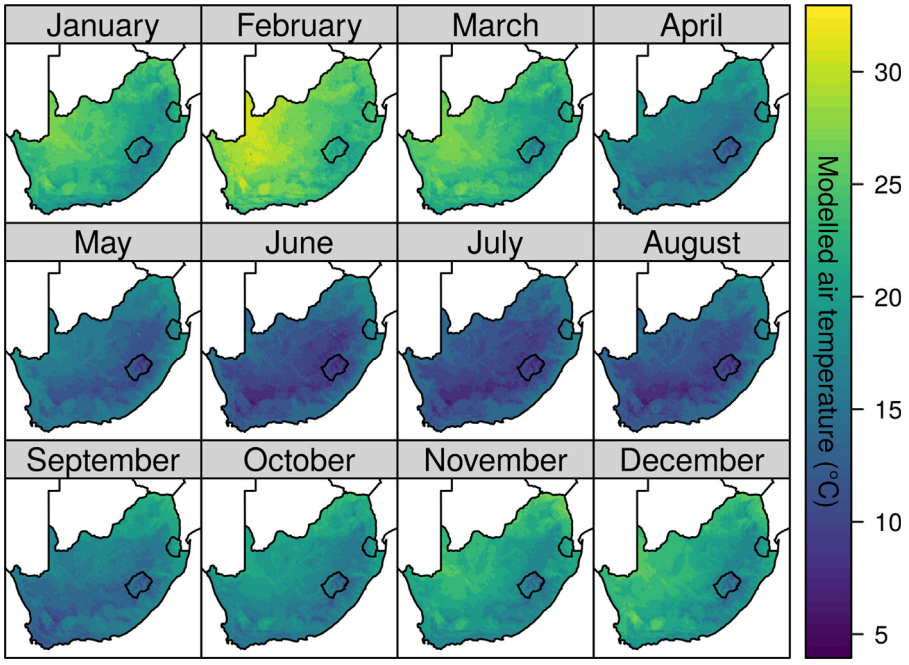

Fig. 5. Exemplary predicted air temperature for the year 2013 represented as monthly means.

spatio-temporal resolution. However, it comes with the limitation that predictions are only possible in non-cloudy conditions. Future improvements might therefore be on appropriate gap-filling techniques in order to provide seamless T_{air} data.

5. Conclusions and data availability

In this study MSG SEVIRI data were used in a machine learning approach to create a gridded and hourly T_{air} data set for entire South Africa. The validation results highlight the high potential of this approach. The validation was performed using spatially and temporally independent data, hence the performance estimates can be considered to be very robust. The study underlines the high potential of MSG SEVIRI to estimate T_{air} in a high temporal and moderate spatial resolution. The resulting data set allows to delineate diurnal variations and can be regarded as a baseline product for climatological analysis on patterns and trends or as a baseline parameter for ecological studies. Limitations of the dataset are (as for all satellite-derived temperature data) the restriction to clear-sky conditions. Also, machine learning strategies such as the used Random Forest algorithm are not able to make extrapolations beyond the range of observed air temperature. In view to climate change and an expected increase of air temperatures, a

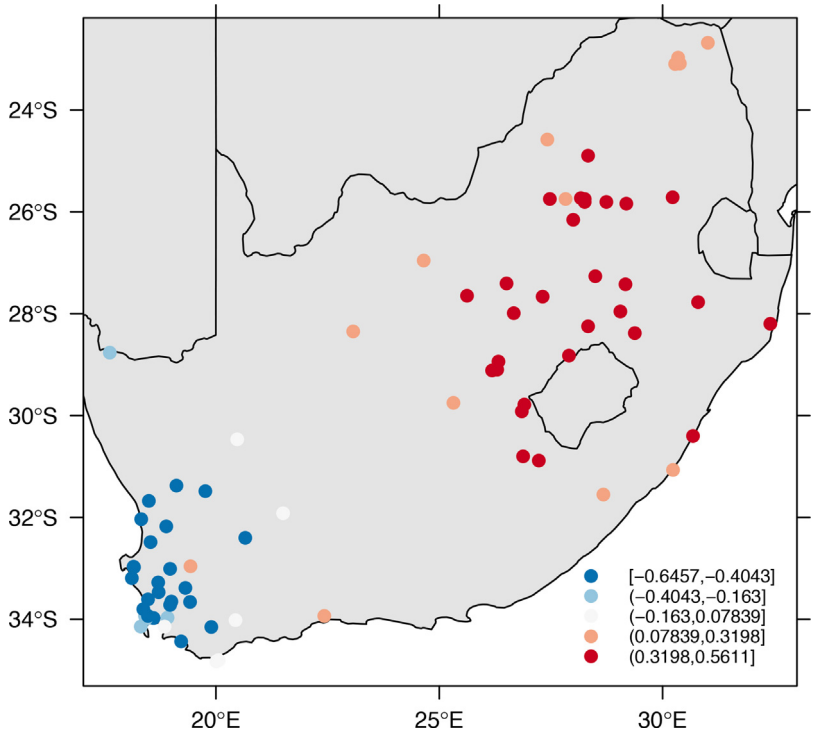

Fig. 6. Spatial representation of the correlation coefficient between the NDVI and air temperature.

re-training of the model will be required if temperatures increase beyond the range that could be observed within the current training period. The hourly gridded air temperature data between 2010 and 2014 can be obtained via https://doi.org/10.1594/PANGAEA.894287. An extension of the data set beyond 2014 as well as for the time between 2004 and 2010 will be made available shortly. The code that led to the creation of the T_{air} data set is available on https://github.com/environmentalinformatics-marburg/Tair4SouthAfrica.

Acknowledgments

This work was funded by the Federal Ministry of Education and Research (BMBF) within the IDESSA project (grant no. 01LL1301) which is part of the SPACES-program (Science Partnership for the Assessment of Complex Earth System processes). We are grateful to the South African Weather Service for providing us with the temperature data from their weather stations. The cloud masking was done by using Level-2 data of the CLAAS-2 data record provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF).

References

Aminou, D.M.A., Jacquet, B., Pasternak, F., 1997. Characteristics of the Meteosat Second Generation (MSG) radiometer/imager: SEVIRI. Proceedings of SPIE: Sensors, Systems, and Next-Generation Satellites, vol. 3221 19–31.

Appelhans, T., Mwangomo, E., Hardy, D.R., Hemp, A., Nauss, T., 2015. Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. Spat. Stat. 14 (Pt A), 91–113. https://doi.org/10.1016/j.spasta.2015.05.008.

Benali, A., Carvalho, A., Nunes, J., Carvalhais, N., Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ. 124, 108–121.

Benas, N., Finkensieper, S., Stengel, M., van Zadelhoff, G.-J., Hanschmann, T., Hollmann, R., Meirink, J.F., 2017. The MSG-SEVIRI-based cloud property data record CLAAS-2. Earth Syst. Sci. Data 9, 415–434. https://doi.org/10.5194/essd-9-415-2017.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Didan, K., 2015a. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS LP DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.006.

Didan, K., 2015b. MYD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MYD13Q1.006.

Eiselt, K.-U., Kaspar, F., Mölg, T., Krähenmann, S., Posada, R., Riede, J.O., 2017. Evaluation of gridding procedures for air temperature over Southern Africa. Adv. Sci. Res. 14, 163–173. https://doi.org/10.5194/asr-14-163-2017.

Elith, J., Graham, C.H., 2009. Do they? how do they? why do they differ? on finding reasons for differing performances of species distribution models. Ecography 32, 66–77. https://doi.org/10.1111/j.1600-0587.2008.05505.x.

EUMETSAT, 2010. High Rate SEVIRI Level 1.5 Image Data – MSG – 0 Degree. (last access: 13 July 2015). https://navigator.eumetsat.int/product/EO:EUM:DAT:MSG:HRSEVIRI.

EUMETSAT, 2012a. Conversion from Radiances to Reflectances for SEVIRI Warm Channels.

EUMETSAT, 2012b. The Conversion from Effective Radiances to Equivalent Brightness Temperatures.

Finkensieper, S., Meirink, J.-F., van Zadelhoff, G.-J., Hanschmann, T., Benas, N., Stengel, M., Fuchs, P., Hollmann, R., Werscheck, M., 2016. CLAAS-2: CM SAF CLoud property dAtAset using SEVIRI – Edition 2. Technical Report Satellite Application Facility on Climate Monitoring.

Giannakos, A., Feidas, H., 2013. Classification of convective and stratiform rain based on the spectral and textural features of Meteosat Second Generation infrared data. Theor. Appl. Climatol. 113, 495–510.

Good, E., 2015. Daily minimum and maximum surface air temperatures from geostationary satellite data. J. Geophys. Res. Atmos. 120, 2306–2324. https://doi.org/10.1002/2014JD022438.

Hengl, T., Heuvelink, G.B.M., Perčec Tadić, M., Pebesma, E.J., 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. Theor. Appl. Climatol. 107, 265–277.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Hooker, J., Duveiller, G., Cescatti, A., 2018. A global dataset of air temperature derived from satellite remote sensing and weather stations. Sci. Data 5, 180246. https://doi.org/10.1038/sdata.2018.246.

Huang, F., Ma, W., Wang, B., Hu, Z., Ma, Y., Sun, G., Xie, Z., Lin, Y., 2017a. Air temperature estimation with MODIS data over the Northern Tibetan Plateau. Adv. Atmos. Sci. 34, 650–662. https://doi.org/10.1007/s00376-016-6152-5.

Huang, W., Li, J., Guo, Q., Mansaray, L.R., Li, X., Huang, J., 2017b. A satellite-derived climatological analysis of urban heat island over shanghai during 2000–2013. Remote Sens. 9. https://doi.org/10.3390/rs9070641.

Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M., Ghazanfari, S., 2017. A statistical framework for estimating air temperature using MODIS land surface temperature data. Int. J. Climatol. 37, 1181–1194. https://doi.org/10.1002/joc.4766.

Jarvis, A., Reuter, H., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the Globe Version 4. Available from the CGIAR-CSI SRTM 90m Database, http://srtm.csi.cgiar.org.

Jury, M.R., 2013. Climate trends in southern Africa. S. Afr. J. Sci. 109, 1–11.

Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perčec Tadić, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res. Atmos. 119, 2294–2313.

Kniffka, A., Stengel, M., Hollmann, R., 2014. SEVIRI Cloud Mask Dataset – Edition 1 – 15 minutes Resolution. Satellite Application Facility on Climate Monitoring. EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF).

Kruger, A.C. (Ed.), 2004. Climate of South Africa. Climate Regions. Report No. WS45. South African Weather Service, Pretoria, South Africa.

Kruger, A.C. (Ed.), 2007. Climate of South Africa. Precipitation. Report No. WS47. South African Weather Service, Pretoria, South Africa.

Kruger, A.C., Sekele, S.S., 2013. Trends in extreme temperature indices in South Africa: 1962–2009. Int. J. Climatol. 33, 661–676. https://doi.org/10.1002/joc.3455.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, 1st ed. Springer, New York.

Kühnlein, M., Appelhans, T., Thies, B., Nauss, T., 2014. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning – a random forests-based approach applied to MSG SEVIRI. Remote Sens. Environ. 141, 129–143.

Meyer, H., 2018. CAST: 'Caret' Applications for Spatial-Temporal Models. R Package Version 0.2.0. https://CRAN.R-project.org/package=CAST.

Meyer, H., Drönner, J., Nauss, T., 2017. Satellite-based high-resolution mapping of rainfall over southern Africa. Atmos. Meas. Tech. 10, 2009–2019.

Meyer, H., Katurji, M., Appelhans, T., Müller, M.U., Nauss, T., Roudier, P., Zawar-Reza, P., 2016. Mapping daily air temperature for Antarctica based on MODIS LST. Remote Sens. 8, 732. https://doi.org/10.3390/rs8090732.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Mucina, L., Rutherford, M.C., 2006. The Vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute.

Neteler, M., 2010. Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data. Remote Sens. 2, 333–351.

Niang, I., Ruppel, O., Abdrabo, M., Essel, A., Lennard, C., Padgham, J., Urquhart, P., 2014. Africa. In: Barros, V., Field, C., Dokken, D., Mastrandrea, M., Mach, K., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., White, L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199–1265.

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E., Stisen, S., 2011. Air temperature estimation with MSG-SEVIRI data: calibration and validation of the TVX algorithm for the Iberian Peninsula. Remote Sens. Environ. 115, 107–116. https://doi.org/10.1016/j.rse.2010.08.010.

Noi, P.T., Degener, J., Kappas, M., 2017. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. Remote Sens. 9. https://doi.org/10.3390/rs9050398.

Prihodko, L., Goward, S.N., 1997. Estimation of air temperature from remotely sensed surface observations. Remote Sens. Environ. 60, 335–346. https://doi.org/10.1016/S0034-4257(96)00216-7.

Shi, L., Liu, P., Kloog, I., Lee, M., Kosheleva, A., Schwartz, J., 2016. Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study. Environ. Res. 146, 51–58.

Stengel, M., Kniffka, A., Meirink, J.F., Lockhoff, M., Tan, J., Hollmann, R., 2014. CLAAS: the CM SAF cloud property data set using SEVIRI. Atmos. Chem. Phys. 14, 4297–4311.

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., Eklundh, L., 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sens. Environ. 110, 262–274. https://doi.org/10.1016/j.rse.2007.02.025.

Vadász, V., 1994. On the relationship between surface temperature, air temperature and vegetation index. Adv. Space Res. 14, 41–44. https://doi.org/10.1016/0273-1177(94)90190-2.

Vancutsem, C., Ceccato, P., Dinku, T., Connor, S.J., 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sens. Environ. 114, 449–465.

Xu, Y., Knudby, A., Shen, Y., Liu, Y., 2018. Mapping monthly air temperature in the Tibetan Plateau from MODIS data based on machine learning methods. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 345–354. https://doi.org/10.1109/JSTARS.2017.2787191.

Zhu, W., Lű, A., Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sens. Environ. 130, 62–73.

Ziervogel, G., New, M., Archer van Garderen, E., Midgley, G., Taylor, A., Hamann, R., Stuart-Hill, S., Myers, J., Warburton, M., 2014. Climate change impacts and adaptation in South Africa. Wiley Interdiscip. Rev. Clim. Change 5, 605–620. https://doi.org/10.1002/wcc.295.